JCM
Journals.ASM.org

Global Phylogeny of *Shigella sonnei* Strains from Limited Single Nucleotide Polymorphisms (SNPs) and Development of a Rapid and Cost-Effective SNP-Typing Scheme for Strain Identification by High-Resolution Melting Analysis

Vartul Sangal,[a] Kathryn E. Holt,[b] Jianfeng Yuan,[a] Derek J. Brown,[c] Ingrid Filliol-Toutain,[d] François-Xavier Weill,[d] Dong-Wook Kim,[e] Wanderley Dias da Silveira,[f] Derek Pickard,[g] Nicholas R. Thomson,[g] Julian Parkhill,[g] Jun Yu[a]

Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, United Kingdom[a]; Department of Microbiology and Immunology, University of Melbourne, Melbourne, Australia[b]; Scottish Salmonella, Shigella & Clostridium difficile Reference Laboratory, Stobhill Hospital, Glasgow, United Kingdom[c]; French National Reference Centre for Escherichia coli and Shigella, Institut Pasteur, Paris, France[d]; Hanyang University, Ansan, South Korea[e]; Biology Institute, University of Campinas, Campinas, Brazil[f]; Wellcome Trust Sanger Institute, Cambridge, United Kingdom[g]

The current *Shigella sonnei* pandemic involves geographically associated, multidrug-resistant clones. This study has demonstrated that *S. sonnei* phylogeny can be accurately defined with limited single nucleotide polymorphisms (SNPs). By typing 6 informative SNPs using a high-resolution melting (HRM) assay, major *S. sonnei* lineages/sublineages can be identified as defined by whole-genome variation.

Shigella sonnei has emerged as a major global cause of shigellosis (1–5). There are 4 distinct *S. sonnei* lineages, and the current global burden involves several geographically associated, multidrug-resistant clones that belong to a single lineage (1). A cost-effective and reliable typing scheme is urgently needed to rapidly identify these clones for active surveillance and epidemiological investigations.

A typing scheme involving 24 single nucleotide polymorphisms (SNPs) has been developed for genotyping *Shigella* serogroups; this system has discriminatory resolution for other *Shigella* serotypes but not for *S. sonnei* (6). Therefore, we focused on the 10,111 SNPs discovered by the genome sequencing of a global collection (1), in order to identify informative SNPs. We selected 97 informative SNPs which were evenly distributed around the *S. sonnei* chromosome (see Table S1 in the supplemental material). We extracted, *in silico*, alleles for these loci from the genome sequences of *S. sonnei* strains Ss046 (NC_007384.1) and 53G (HE616528) and 99 representative isolates (ERP000182) described by Holt et al. (1) (see Table S2 in the supplemental material). A minimal spanning tree was generated using Bionumerics 5.1 (Applied Biomaths, Belgium) by grouping haplotypes that shared at least 92/97 loci. The overall structure of the tree was nearly identical to that produced by the 10,111 SNPs (see Fig. S1 in the supplemental material); all the isolates were resolved in five clusters (Fig. 1A). Most of the 99 isolates analyzed previously fell into lineages I, II, and III (gray nodes in Fig. 1A). The minor differences were that (i) strain 259, which stood alone as lineage IV, was grouped with lineage I isolates because only 3 of 97 selected SNPs separated these lineages and (ii) nine isolates were separated from lineages I to III as singletons due to differences at more than five loci. However, these singletons shared more alleles with the isolates in their corresponding lineages than isolates from other lineages. Hence, the 97 SNPs were phylogenetically informative and can reliably discriminate between the known *S. sonnei* lineages.

To test the phylogenetic robustness of these 97 SNPs, we analyzed 68 novel isolates using high-resolution melting (HRM) (see Table S2 in the supplemental material). These 68 isolates were collected between 1991 and 2007, including 13 from Africa, 19 from Asia, 2 from the Caribbean, 2 from Europe, 4 from North America, and 10 from South America. The origins of 18 isolates were not recorded. For each isolate tested, 100- to 150-bp fragments were amplified for each SNP locus, followed by HRM according to the manufacturer's instructions (Qiagen). SNPs were called against the reference Ss046 strain (7) using Rotor-Gene 6000 Series Software 1.7 with ≤30% confidence. Loci with ≥80% confidence were called as reference alleles, and variants with 31 to 79% confidence were considered ambiguous and were resolved by sequencing.

Of the 68 isolates analyzed by HRM, 10 belonged to lineage II and 58 belonged to lineage III (Fig. 1A). The new lineage II isolates included two from Brazil (1997, 1999), one from the Dominican Republic (2006), and one from Bali (2001); the remainder had unknown geographical provenance and were isolated between 1991 and 2007. This demonstrates that lineage II had reached South America by the late 1990s and confirmed its presence in Asia. Four isolates, three from Mexico (2001 to 2003) and one of unknown origin (2002), were clustered in IIIa, which has been previously shown to be present in Europe and Central America/the Caribbean. This indicated that sublineage IIIa had also dispersed to North American countries. A total of 28 novel isolates belonged to sublineage IIIb, most of which originated in Brazil (1997 to 2002) or the Indian subcontinent (Pakistan, India, and Nepal; 2000 to 2008). This is consistent with previous observa-

Received 22 August 2012 Returned for modification 3 October 2012
Accepted 23 October 2012
Published ahead of print 31 October 2012
Address correspondence to Jun Yu, jun.yu@strath.ac.uk.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.02238-12.

doi:10.1128/JCM.02238-12

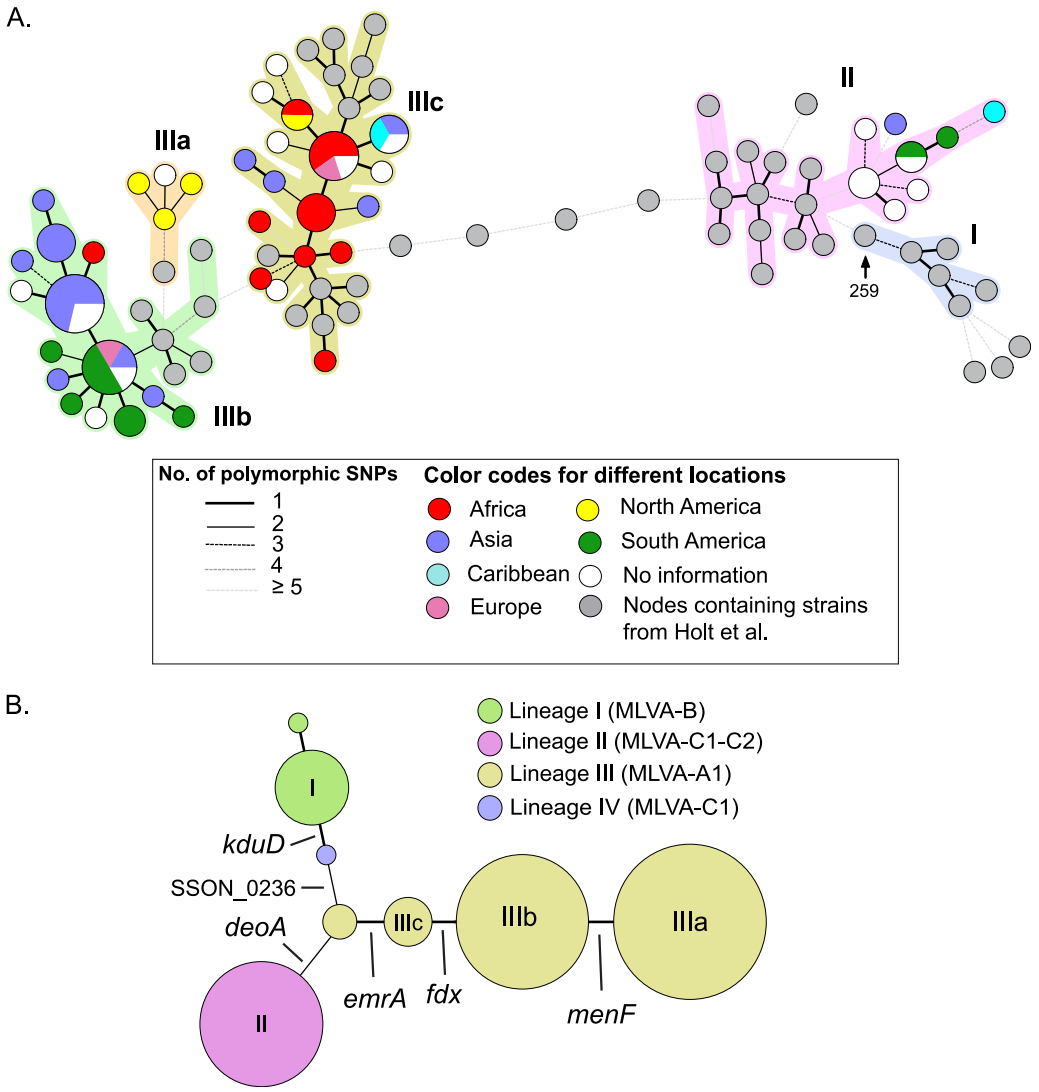

FIG 1 (A) A minimal spanning tree of 169 *S. sonnei* isolates using the data from 97 chromosomal SNPs. Haplotypes were clustered into groups by sharing at least 92 alleles. Nodes including isolates described by Holt et al. (1) are shown in gray and do not reflect the number of isolates. Nodes containing 68 isolates that were HRM typed are displayed as pie charts whose sizes correspond to the numbers of isolates. Distances between haplotypes (nodes) reflect the numbers of polymorphic SNPs and are shown in log scale. (B) Major *S. sonnei* clones defined by a minimal set of six SNPs. Lineages are labeled as identified from the whole genomic variation presented by Holt et al. (1) and their corresponding MLVA groups as described by Filliol-Toutain et al. (8).

tions of IIIb in Pakistan and French Guyana. A total of 26 of the 68 isolates belonged to sublineage IIIc, which included most of the isolates from Africa (Malawi, Gambia, Nigeria, Morocco, Egypt, and Tunisia; 2000 to 2006), five from Asia (China, Uzbekistan, and Syria; 2002 to 2007), and one each from the Caribbean (Cuba; 2005), Europe (Turkey; 2006), and North America (Mexico; 2005). The geographic origins of the remaining six IIIc isolates (2000 to 2007) were unknown. These results were also consistent with the previous report (1) that IIIc was globally present, with the majority of isolates from Africa and Asia.

The phylogeny produced by the 97 SNPs was in agreement with the structure produced by multilocus variable-number tandem-repeat analysis (MLVA), which produced 4 main clones: A1, B, C1, and C2 (8). Isolates from MLVA group A1 were distributed among sublineages IIIa to IIIc, while all MLVA group B isolates belonged to lineage I and isolates of MLVA groups C1 and C2 belonged to lineage II (see Fig. S2 in the supplemental material).

Recombination has been shown to be rare within *S. sonnei* (1). We found a strong linkage disequilibrium between 97 SNPs (index of association $[I_A] = 47.13$, $P < 0.01$; standardized measure of I_A: $r_D = 0.50$, $P < 0.01$) using MultiLocus 1.3 (9). Due to this nonrandom association between 97 SNPs, we further identified six SNPs that can define the main lineages/sublineages (see Fig. S1 in the supplemental material). Four of these 6 SNPs (*kduD*, *deoA*, *SSON_0236*, and *emrA*) resolved isolates into lineages I, II, and IV and collectively into lineage III. Two additional SNPs (*fdx* and *menF*) further resolved lineage III into sublineages IIIa, IIIb, and IIIc (Fig. 1B; see Fig. S1). Furthermore, an SNP within the *gyrA* gene was present in all isolates that were resistant to quinolones or that showed reduced susceptibility to fluoroquinolones. Therefore, this SNP can be tested to identify quinolone-resistant isolates (see Table S1 in the supplemental material).

Several molecular typing schemes, including multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE),

and MLVA, have been applied to *S. sonnei* for evolutionary and epidemiological studies. MLST lacks discrimination, as all isolates are assigned to a single clonal complex (10). PFGE and MLVA offer higher resolution (8, 11, 12), but these methods are more labor-intensive, require special expertise, and are difficult to interpret. In contrast, HRM-based SNP typing is highly sensitive, rapid, and cost-effective and is therefore increasingly used for bacterial analysis (13–15). HRM data analysis does not require expert training, and it can be easily implemented in reference laboratories across the globe provided an initial investment in an HRM facility is made. Whole-genome sequencing (WGS) indexes the entire genome variation and is therefore suitable for studying details of the population structure. However, WGS of a strain still costs more than six PCRs and poses technical challenges in handling and analyzing the enormous amounts of data required for routine diagnostics. Moreover, this set of six SNPs has been derived with a highly robust phylogeny based on genome-wide variation in a diverse global collection, and an absence of recombination in *S. sonnei* (1) provides them long-term stability as molecular markers for the rapid detection of major *S. sonnei* clones.

ACKNOWLEDGMENTS

We thank Jiang Yang (Institute of Pathogen Biology, Beijing, China) for useful discussion.

This study was financed by an MRC grant to J. Yu (G0800173). K. E. Holt was supported by a fellowship from the NHMRC of Australia (number 628930). D.-W. Kim is supported by the governments of South Korea, Sweden, and Kuwait. N. R. Thomson, D. Pickard, and J. Parkhill were financed by The Wellcome Trust (grant number 0689).

REFERENCES

1. **Holt KE, Baker S, Weill FX, Holmes EC, Kitchen A, Yu J, Sangal V, Brown DJ, Coia JE, Kim DW, Choi SY, Kim SH, da Silveira WD, Pickard DJ, Farrar JJ, Parkhill J, Dougan G, Thomson NR.** 2012. *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat. Genet. **44**:1056–1059.
2. **Ranjbar R, Soltan Dallal MM, Talebi M, Pourshafie MR.** 2008. Increased isolation and characterization of *Shigella sonnei* obtained from hospitalized children in Tehran, Iran. J. Health Popul. Nutr. **26**:426–430.
3. **Seol SY, Kim YT, Jeong YS, Oh JY, Kang HY, Moon DC, Kim J, Lee YC, Cho DT, Lee JC.** 2006. Molecular characterization of antimicrobial resistance in *Shigella sonnei* isolates in Korea. J. Med. Microbiol. **55**:871–877.
4. **von Seidlein L, Kim DR, Ali M, Lee H, Wang X, Thiem VD, Canh do G, Chaicumpa W, Agtini MD, Hossain A, Bhutta ZA, Mason C, Sethabutr O, Talukder K, Nair GB, Deen JL, Kotloff K, Clemens J.** 2006. A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. PLoS Med. **3**:e353. doi:10.1371/journal.pmed.0030353.
5. **Wei HL, Wang YW, Li CC, Tung SK, Chiou CS.** 2007. Epidemiology and evolution of genotype and antimicrobial resistance of an imported *Shigella sonnei* clone circulating in central Taiwan. Diagn. Microbiol. Infect. Dis. **58**:469–475.
6. **Hayford AE, Mammel MK, Lacher DW, Brown EW.** 2011. Single nucleotide polymorphism (SNP)-based differentiation of *Shigella* isolates by pyrosequencing. Infect. Genet. Evol. **11**:1761–1768.
7. **Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, Tang X, Wang J, Xiong Z, Dong J, Xue Y, Zhu Y, Xu X, Sun L, Chen S, Nie H, Peng J, Xu J, Wang Y, Yuan Z, Wen Y, Yao Z, Shen Y, Qiang B, Hou Y, Yu J, Jin Q.** 2005. Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. Nucleic Acids Res. **33**:6445–6458.
8. **Filliol-Toutain I, Chiou CS, Mammina C, Gerner-Smidt P, Thong KL, Phung DC, Pichel M, Ranjbar R, Sow AG, Cooper K, Ribot E, Binsztein N, Liang SY.** 2011. Global distribution of *Shigella sonnei* clones. Emerg. Infect. Dis. **17**:1910–1912.
9. **Agapow P-M, Burt A.** 2001. Indices of multilocus linkage disequilibrium. Mol. Ecol. Notes **1**:101–102.
10. **Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol. Microbiol. **60**:1136–1151.
11. **Chiou CS, Watanabe H, Wang YW, Wang WL, Terajima J, Thong KL, Phung DC, Tung SK.** 2009. Utility of multilocus variable-number tandem-repeat analysis as a molecular tool for phylogenetic analysis of *Shigella sonnei*. J. Clin. Microbiol. **47**:1149–1154.
12. **Kim KS, Oh JY, Jeong YW, Cho JW, Park JC, Cho DT, Lee JC.** 2002. Epidemiological typing and characterization of dfr genes of *Shigella sonnei* isolates in Korea during the last two decades. J. Microbiol. Biotechnol. **12**:106.
13. **Bratchikov M, Mauricas M.** 2011. Development of a multiple-run high-resolution melting assay for *Salmonella* spp. genotyping HRM application for *Salmonella* spp. subtyping. Diagn. Microbiol. Infect. Dis. **71**:192–200.
14. **Grando D, Said MM, Mayall BC, Gurtler V.** 2012. High resolution melt analysis to track infections due to ribotype 027 *Clostridium difficile*. J. Microbiol. Methods **89**:87–94.
15. **Lilliebridge RA, Tong SY, Giffard PM, Holt DC.** 2011. The utility of high-resolution melting analysis of SNP nucleated PCR amplicons—an MLST based *Staphylococcus aureus* typing scheme. PLoS One **6**:e19749. doi:10.1371/journal.pone.0019749.